Bearing Fault Detection in Conveyor Belt Drums Using Machine Learning.

Victor Afonso Bauler[1], Júlio A. Cordioli[2], Danilo Silva[3], and Danilo Braga[4]

[1,2,3] *Federal University of Santa Catarina, Florianópolis, Brazil*
victor.bauler@lva.ufsc.br, julio.cordioli@ufsc.br, danilo.silva@ufsc.br

[4] *Dynamox SA, Florianópolis, Brazil*
danilo@dynamox.net

ABSTRACT

In recent years, the application of machine learning techniques in condition monitoring has significantly advanced the precision and efficiency of fault detection processes. In particular, detecting bearing faults in conveyor belt drums is critical in the mining industry for maintaining operational reliability and productivity. This paper presents a case study using vibration signals and diagnostic reports provided by the company Dynamox. After meticulous data cleaning, preprocessing, and feature extraction employing advanced signal processing techniques and statistical features, several machine learning models were trained, optimized and evaluated, with the best models providing very promising results.

1. INTRODUCTION

The evolution of technology and the rise of Industry 4.0 have provided new opportunities and challenges in the field of industrial maintenance. The integration of the Internet of Things (IoT), cloud computing, and artificial intelligence has redefined traditional approaches to machine monitoring and maintenance . In this context, machine condition monitoring (CM), and more specifically, vibration monitoring, has emerged as one of the most effective techniques, allowing maintenance teams to identify potential problems in advance and plan interventions strategically, minimizing impacts on production and operation (Randall, 2021).

In the mining industry, condition monitoring is confronted with unique challenges due to the extreme environmental conditions and the broad geographic distribution of assets, which can extend over several kilometers (Zimroz & Król, 2015). Due to the risks and complexities involved in monitoring all equipment in the field, techniques that leverage IoT sensors for condition monitoring offer significant advantages. These include the ability to continuously collect data, which then can be analyzed using algorithms for fault detection. This approach provides vibration analysts with powerful tools to assess and maintain equipment health.

The conveyor belt plays a crucial role in the transportation systems of the industry. It is a mechanical device used for the continuous movement of materials over short or long distances. Its main components include the motor, rollers, drums, belt, loading chute, and pulleys (Zimroz & Król, 2015). Conveyor belt drums, particularly their bearings, stand out due to their susceptibility to faults (Bortnowski, Kawalec, Król, & Ozdoba, 2022). These faults, if not detected and rectified timely, can lead to substantial operational downtimes, affecting the overall productivity and economic efficiency of mining activities.

This study investigates the application of machine learning techniques for condition monitoring of conveyor belt drums, with a focus on detecting bearing faults. We evaluate the performance of four distinct models: Logistic Regression, Support Vector Machines (SVM), RandomForest, and XGBoost. Deep learning models are not considered in this analysis, as our processing pipeline already includes a feature extraction step, making traditional machine learning models more suitable for effectively utilizing the pre-processed data.

The methodology section outlines the comprehensive steps undertaken in this research, including data acquisition, feature engineering, the division of data into training and testing sets, a detailed description of the models, and the process of hyperparameter optimization. Subsequently, the results section presents the outcomes derived from each model based on the methodology employed. In conclusion, this paper highlights the key findings of the study, underscoring its contributions to the field and suggesting directions for future research.

Figure 1. Belt conveyors in iron ore yard (Carvalho, 2022).

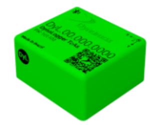

Figure 2. DynaLogger TcAs.

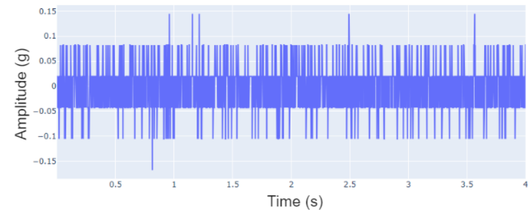

Figure 3. Idle machine signal.

2. DATA ACQUISITION

The dataset provided by the partner company consists of two different data sources: vibration signals and diagnostic reports made by vibration analysts. Data was collected over a two-year period, with daily measurements. The reports, prepared by the company's vibration analysts, include information on the identified failure modes, the report creation date, the diagnosis, and other details pertaining to each monitored component; all the client data was anonymized. The analysts conducted the report assessments at approximately two-week intervals for each measurement point.

The dataset employed in this study encompasses 61 conveyor belts, with an illustrative example depicted in Figure 1. It includes 550 measurement points focused on drum bearings. To enhance data quality and relevance, we initially apply filters to the vibration data to exclude periods when machines were idle, as well as to remove any instances of clipping and discrepancies in metadata settings. Subsequently, we align the closest measurement to the date of each diagnostic report for every measurement point. This method ensures that our models are trained and tested using only the waveform data that is nearest in time to each report.

Vibration data are captured using DynaLoggers, wireless MEMS-type sensors produced by Dynamox, equipped with temperature and triaxial vibration measurement capabilities. These devices are securely attached to the equipment, either by gluing or screwing onto the casing, and collect data at predefined intervals. Data is stored in the sensor's internal memory until it is transmitted to a cloud-hosted Web Platform by an external collector, such as a mobile app or gateway. This setup enables remote asset monitoring, eliminating the need for physical data collection trips and thereby reducing associated risks and saving time for analysts who benefit from continuous access to updated data. Figure 2 displays the DynaLogger TcAs model, the specific model employed in this study.

DynaLogger TcAs offer two distinct types of monitoring: telemetry and spectral. Telemetry monitoring encompasses predefined metrics derived from the signal, such as RMS, peak, peak to peak, crest factor, skewness, kurtosis, and contact temperature. In contrast, spectral monitoring involves longer duration acquisitions, providing access to raw vibration data. This enables the use of various analysis techniques, including spectrum analysis, waveform (linear, circular, and orbital) analysis, frequency filters, cepstrum, spectral envelope (demodulation), autocorrelation, and multi-metrics analysis. For this study, we exclusively utilized data from spectral monitoring, which was sampled at a rate of 2048 Hz over a 4.0-second duration.

The idle machine filter operates by examining the number of unique values in each signal, taking into account that the sensor's sampling resolution is 8 bits. This approach is tailored to the dynamic range set by the client; if the amplitude level of a signal is low, the sensor is likely to encounter a quantization issue due to this limited resolution. An example of such a signal, illustrating the impact of quantization at low amplitude levels, is presented in Figure 3.

The method for identifying and discarding signals with clipping hinges on analyzing the histogram of each signal. This analysis focuses on whether a significant concentration of points exists within a range exceeding 70% of the signal's dynamic range—a threshold established experimentally. If the concentration of points in this range is notably high, the signal is considered to be clipped and is subsequently discarded. This approach and its rationale are visually illustrated and detailed in Figure 4.

The final filtering process applies to both measurements and analysts' reports to address inconsistencies in metadata. An example of such inconsistency is a discrepancy between the

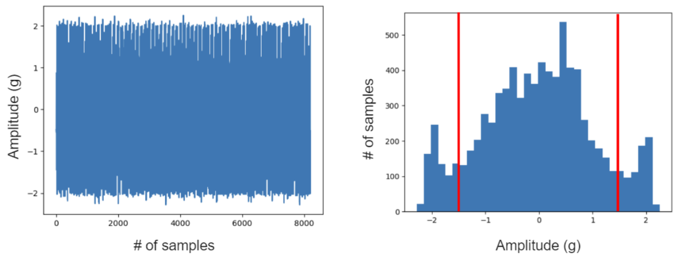

Figure 4. Clipped signal and its corresponding histogram.

Table 1. Failures distribution.

Failure	Number of Samples
Bearing Wear	2047
Stiffness Loss	394
Lack of Lubricant	42
Misalignment	20
Others	33
No Failure	4466

signal length and the product of the filled duration and the sampling frequency. In the case of reports, the most common issues include the absence of bearing information and other missing details. After pairing each report with its closest measurements, we have compiled a dataset of 7002 paired measurements and reports. This dataset will serve as the foundation for the analysis in this study. Table 1 presents a comprehensive breakdown of all types of failures identified within this dataset.

Our objective is to specifically identify bearing failures, hence we will encode all instances of bearing wear failures with 1 and categorize all other types of failures with 0. This binary classification will enable us to effectively train our machine learning models to recognize bearing failures.

3. FEATURE EXTRACTION

In this study, our focus was on feature extraction from a single axis, designated in our preprocessing steps as the "main axis." This axis is identified by having the highest acceleration RMS value. We adopted this approach due to the unavailability of data from all three axes in every measurement.The feature set is organized into categories based on the time and frequency domains, encompassing acceleration, velocity, and envelope features. In the time domain, the statistical metrics employed include root mean square (RMS), peak-to-peak, kurtosis, skewness, crest factor, and shape factor. Envelope features are derived through envelope analysis, which involves (i) applying a passband filter to a specified region of interest; (ii) extracting the signal's envelope, in this case, using the Hilbert transform; and (iii) computing the spectrum of the envelope.

The estimated machine rotation speed is input into the Dynamox platform by clients, along with the registration of rolling bearings. This enables access to the values of BPFO (Ball Pass Frequency of the Outer race), BPFI (Ball Pass Frequency of the Inner race), BSF (Ball Spin Frequency), and FTF (Fundamental Train Frequency) for each measurement point, facilitating targeted feature extraction and analysis.

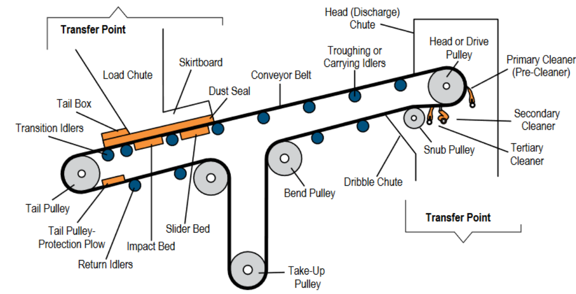

Figure 5. Basic Mechanical Components of a Conveyor (Swinderman, 2009).

All features, both in the time and frequency domains, were calculated globally and for different frequency ranges, depending on the specific metric. The frequency intervals and harmonics used for the extraction of specialized features are listed in Table 2.

Another important aspect of our dataset includes metadata detailing the location of the belt drum. This metadata encompasses a range of specific locations such as bend pulley, drive pulley, take-up pulley, and others, each offering unique insights into the operational context of the conveyor belt system. Understanding the spatial aspect of where faults occur can significantly enhance our analysis, allowing us to pinpoint more precisely the areas prone to issues and potentially revealing patterns related to specific locations. Figure 5 illustrates the basic mechanical components of the conveyor belt, showing some of the cited locations.

4. MACHINE LEARNING METHODS

The machine learning models utilized in this study were implemented using scikit-learn and xgboost, widely-used machine learning libraries in Python (Pedregosa et al., 2011), (Chen & Guestrin, 2016). The subsequent subsection presents the specific models used, along with details of the adopted hyperparameter optimization algorithm and the train test split.

Logistic Regression, alternatively known as Logit Regression, is a regression algorithm widely used for classification tasks. It operates similarly to Linear Regression by determining a weighted sum of the input attributes, plus a constant bias. However, Logistic Regression diverges from Linear Regression in its application of the logistic (sigmoid) function to this sum. This sigmoid function modifies the output to fall within a 0 to 1 range, thereby producing a probability out-

Table 2. Features extracted from the signals.

Representation	Metrics	Frequency Range (Hz)/Harmonics(X)
Acceleration	RMS	Global, [5, 200], [5, 400], [5, 1000]
	Kurtosis	
	Skewness	
	Crest Factor	
	Shape Factor	
Velocity	RMS	Global, [5, 50], [50, 110], [5, 200], [5, 400], [500, 1000]
	Peak to Peak	
Envelope	RMS	[5, 200], [5, 400], [50, 200], [200, 400] 1X, 2X, 3X, 4X, 5X
	BPFO	
	BPFI	
	BSF	

come. This mechanism makes Logistic Regression particularly suited for binary classification tasks, transforming raw linear outputs into interpretable probabilities (Géron, 2022).

SVM is a powerful algorithm for classification tasks that aims to find an optimal hyperplane for separating different classes (Watt, Katsaggelos, & Borhani, 2016). Diverging from the Logistic Regression, SVM adopts a distinct cost function and optimization method. This approach transforms the model into non-parametric, capable of identifying non-linear relationships through the application of a kernel function. Random Forest is an ensemble learning method that combines multiple decision trees to make predictions in classification and regression tasks (Géron, 2022). It utilizes bagging, which involves creating subsets of the training data and constructing decision trees independently on each subset. The final prediction is achieved by leveraging soft voting, which aggregates the predictions from each decision tree by computing the average probability assigned to each class.

XGBoost stands as a sophisticated algorithm in machine learning, leveraging a mix of weak learners, like low-depth decision trees, to construct a powerful predictive model. It utilizes a boosting approach, sequentially incorporating new models into the ensemble with an emphasis on correcting previously incorrectly classified instances (Chen & Guestrin, 2016). Effective evaluation of predictive models is a crucial step in their development. A fundamental part of this evaluation is the strategy of dividing the data into training and test sets, which must be carefully planned to ensure the validity and robustness of the results obtained. Data leakage can lead to an overly optimistic assessment of the model, which does not reflect its true generalization capability. In the context of fault detection in bearings, many studies neglect this step, as reported by (Hendriks, Dumond, & Knox, 2022).

In this study, we will split the dataset by grouping measurement points that are physically proximate to each other in the real world, ensuring these groups are allocated to the same set (either train or test). Subsequently, we will randomly select 20% of these groups to construct our test set.

Hyperparameter tuning is essential in the development of ma-

Table 3. Hyperparameters and their search space for each model.

Model	Hyperparameter	Search Space
Logistic Regression	C	$10^{-4} : 10^{4}$
SVM	C	$10^{-4} : 10^{4}$
	γ	$10^{-4} : 10^{4}$
	Kernel	RBF
Random Forest	n_estimators	$100, 400, 800$
	max_depth	$3 : 31$
	CCP α	$10^{-3} : 1$
	min_samples_split	$2 : 101$
	min_samples_leaf	$1 : 101$
XGBoost	n_estimators	$100, 400, 800$
	max_depth	$3 : 31$
	η	$0 : 0.3$
	γ	$10^{-3} : 10^{3}$
	subsample	$0.5 : 1$
	colsample_bytree	$0.5 : 1$

chine learning models, aiming to find the best hyperparameter settings to optimize model performance. Hyperparameters, which are adjustable parameters that control the model's learning process, include factors such as learning rate, regularization strength, and the number of estimators. Choosing the right hyperparameters can greatly enhance the model's accuracy and its ability to generalize from training data. To measure validation performance and optimize hyperparameters, this study utilizes group k-fold cross-validation exclusively on the training dataset. This method partitions the data into groups to rigorously assess model performance, avoiding the biases associated with single random data splits, we choose . Table 3 presents the hyperparameters tested and their corresponding search spaces for each model. A random search with 100 iterations was employed to explore the hyperparameter space.

As shown in Table 1, the dataset exhibits a substantial imbalance between the classes of defects. In this context, metrics that are not affected by the prevalence of the classes were employed, meaning metrics that do not vary according to the proportion of the classes. Therefore, to assess the predictive capability of the models, two metrics were used: ROC AUC (Area Under the Receiver Operating Characteristic Curve) and the false positive rate for a given true positive rate. The ROC AUC, it is a metric that evaluates a classification model's ability to accurately distinguish between positive and negative classes. The ROC curve is constructed by plotting the true positive rate (TPR) against the false positive rate (FPR).

The optimal hyperparameter selection will be determined using the ROC AUC metric during Cross Validation. Following this, the models will undergo retraining utilizing the entire training dataset and will be subsequently assessed on the test set. Within the test set evaluation, our objective is to minimize the FPR while ensuring that the TPR is at least 90

5. RESULTS

This study presents a comprehensive evaluation of four machine learning models: Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost, in the context of binary classification tasks. Table 4 outlines the performance of each model across training, cross-validation, and testing phases, with a focus on the ROC AUC and the False Positive Rate (FPR) at a True Positive Rate (TPR) of 90% or higher.

The results in Table 4 highlight the performance of the evaluated models. Logistic Regression shows strong generalization with a test ROC AUC of 0.93, though with a moderate FPR of 0.23 at a TPR of 90%. SVM, despite achieving a high training ROC AUC (0.95), experienced a performance drop during testing (ROC AUC of 0.88) and displayed the highest FPR (0.50), indicating susceptibility to overfitting. Random Forest maintained robust test performance (ROC AUC of 0.94) with the lowest FPR (0.19), suggesting a good balance between accuracy and false positive control. XGBoost, while achieving perfect performance during training (ROC AUC of 1.00), exhibited a slight decrease in test ROC AUC (0.93) and a higher FPR (0.29).

In the subsequent analysis, a graphical representation of the ROC curves for all four models—Logistic Regression, SVM, Random Forest, and XGBoost—is provided to visually compare their performance. Despite the varying numeric metrics discussed earlier, the ROC curves of these models exhibit a remarkable closeness, as depicted in Figure 6. This proximity in the curves suggests that, from a graphical standpoint, the models' abilities to discriminate between the classes are nearly indistinguishable, with the exception of the SVM.

Further insights into model interpretability and the factors influencing predictions were gained through the application of SHAP (SHapley Additive exPlanations) analysis, particularly on the Random Forest model. The SHAP summary plot visualizes the impact of features on model predictions through three components: the features (y-axis), their SHAP values (x-axis), and the feature values (colorbar). Features are

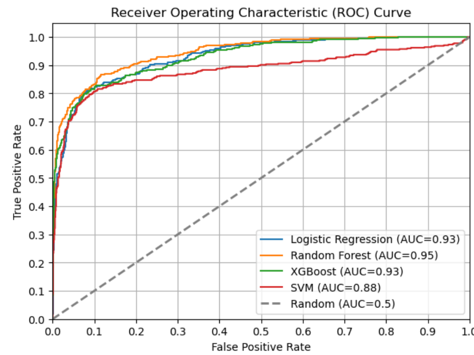

Figure 6. ROC curves.

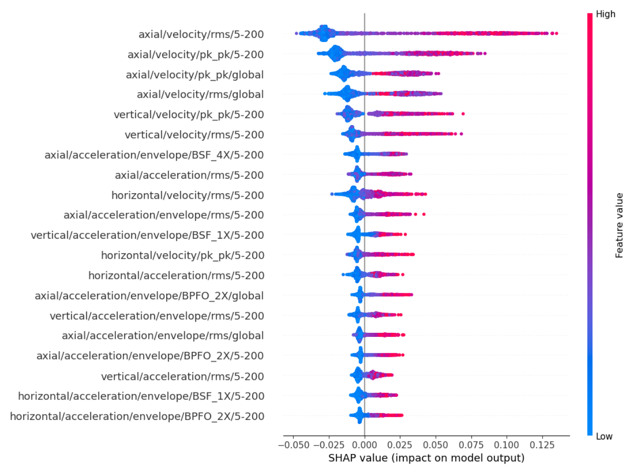

Figure 7. SHAP summary of Random Forest.

ranked on the y-axis, while the x-axis indicates their SHAP values, where positive values drive predictions towards 1 (defect) and negative values towards 0 (healthy). The colorbar denotes feature values, with colder colors (e.g., blue) representing lower values and warmer colors (e.g., red) indicating higher values.

Figure 7 reveals that the majority of significant features originate from envelope analysis, underscoring its critical role, particularly in identifying the predominant harmonic BPFO. Additionally, it emphasizes the prominence of velocity features, especially noting that the top two features are associated with RMS. Notably, the frequency range of 5-200 Hz emerged as the most impactful in the analysis. This can likely be attributed to bearing fault frequencies being directly visible in the FFT, indicative of advanced fault stages.

6. CONCLUSION

This research explored machine learning models to improve condition monitoring for conveyor belt drums, focusing on detecting bearing faults in the mining industry's challenging environment.. Through the meticulous application and

Table 4. Models results.

Model	Train	Cross Validation	Test	
	ROC AUC	ROC AUC	ROC AUC	FPR (TPR $\geq$ 90%)
LogisticRegression	0.87	0.82	0.93	0.23
SVM	0.95	0.84	0.88	0.50
RandomForest	0.93	0.83	0.94	0.19
XGBoost	1.00	0.85	0.93	0.29

evaluation of Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost models, we've explored the potential of these algorithms to accurately classify bearing conditions as either healthy or defective. Our findings, as summarized in Table 4 and further illustrated by the ROC curves on Figure 6.

Random Forest emerged as the best-performing model, achieving the highest test ROC AUC (0.94) and the lowest false positive rate (FPR) of 0.19, demonstrating a strong balance between accuracy and minimizing false alarms. Logistic Regression and XGBoost both performed well, with test ROC AUC values of 0.93, though Logistic Regression had a slightly lower FPR (0.23) compared to XGBoost (0.29). The SVM, despite strong training performance, showed the weakest results during testing with a ROC AUC of 0.88 and the highest FPR of 0.50, indicating issues with overfitting.

The ROC curves (Figure 6) indicate that while most models demonstrated similar class discrimination, Random Forest stood out for its superior overall performance. The SVM curve diverged more significantly due to its relatively lower test performance. In addition, SHAP analysis was applied to the Random Forest model, showing that envelope analysis and velocity features, particularly within the 5-200 Hz frequency range, played a key role in detecting bearing faults, further highlighting the importance of the feature engineering step.

In conclusion, our study underscores the promise of machine learning techniques in advancing condition monitoring practices, particularly within the mining industry's challenging operational context. While each model offers unique advantages, their selection should be tailored to specific monitoring objectives and operational constraints. Future research should continue to refine these models, explore the integration of additional data sources, and further investigate the implications of feature selection on model accuracy and interpretability. Through such efforts, the goal of achieving more reliable, efficient, and cost-effective condition monitoring solutions in industrial settings becomes increasingly attainable.

Moving forward, it's essential to address the subjectivity introduced by some vibration analysts, a factor attributed to the field's high personnel turnover. This variability in the interpretation of data can lead to inconsistent labeling, presenting a significant challenge in training our machine learning models with precision. By acknowledging this aspect, our future endeavors will aim to mitigate the impact of subjective analysis on our dataset.

Furthermore, the development and refinement of these machine learning models are intended to augment the workflow of analysts. By providing robust tools that can accurately identify potential faults in machinery, we're not only aiming to enhance the precision of diagnostics but also to support analysts in making more informed decisions. This augmentation is particularly valuable in a field characterized by frequent changes in personnel, as it contributes to standardizing the analysis process, thereby reducing the likelihood of errors and inconsistencies. Ultimately, these models serve as a complement to the expertise of vibration analysts, leveraging the power of artificial intelligence to enrich the condition monitoring process with deeper insights and greater efficiency.

ACKNOWLEDGMENT

We thank Dynamox S/A, GAMA UFSC and the Laboratory of Vibrations and Acoustics at UFSC, supported by FEESC, for making this work possible through the project "Technologies for the analysis of noise and vibration signals for the detection, prognosis, and correction of failures in machines and structures".

REFERENCES

Bortnowski, P., Kawalec, W., Król, R., & Ozdoba, M. (2022, October). Types and causes of damage to the conveyor belt – review, classification and mutual relations. *Engineering Failure Analysis*, *140*, 106520. doi: 10.1016/j.engfailanal.2022.106520

Carvalho, Y. S. C. d. (2022). *Failure detection in belt conveyor drum bearings using accelerometers and machine learning.* (Unpublished master's thesis). Federal University of Ouro Preto.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (p. 785–794). doi: 10.1145/2939672.2939785

Géron, A. (2022). *Hands-on machine learning with scikit-*

learn, keras, and tensorflow. O'Reilly.

Hendriks, J., Dumond, P., & Knox, D. A. (2022, April). Towards better benchmarking using the cwru bearing fault dataset. *Mechanical Systems and Signal Processing*, *169*, 108732. doi: 10.1016/j.ymssp.2021.108732

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Randall, R. B. (2021). *Vibration-based condition monitoring* (2nd ed.). Wiley.

Swinderman, R. T. (2009). *Foundations: The practical resource for cleaner, safer, more productive dust and material control* (4th ed.).

Watt, J., Katsaggelos, A. K., & Borhani, R. (2016). *Machine learning refined: foundations, algorithms, and applications*. Cambridge: Cambridge university press.

Zimroz, R., & Król, R. (2015, September). Failure analysis of belt conveyor systems for condition monitoring purposes. *Mining Science*, *128*(36), 255–270.

BIOGRAPHIES

Victor A. Bauler is enrolled at the Mechanical Engineering Masters with emphasis in Vibrations and Acoustics at Federal University of Santa Catarina (UFSC), Brazil, where he also graduated in 2023. He has experience with machine learning and is currently doing research in domain adaptation and domain generalization for bearing fault detection.

Prof. Julio A. Cordioli graduated in Mechanical Engineering from Federal University of Santa Catarina (UFSC), Brazil, in 2000 and received his PhD in 2006 also from UFSC, with part of his PhD research developed at the University of Cambridge, UK. He has previously worked for EMBRAER (Brazil) on the design of noise control treatments for aircrafts, and for the ESI Group (USA), as an acoustic scientist on the development of vibroacoustic numerical methods. Since 2011, he has been an Associate Professor at the Mechanical Engineering Department, UFSC. His main research interests include: numerical methods in vibro-acoustics, aircrafts interior and exterior noise, aeroacoustics and vibration monitoring.

Prof. Danilo Silva received the B.Sc. degree from the Federal University of Pernambuco (UFPE), Recife, Brazil, in 2002, the M.Sc. degree from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, in 2005, and the Ph.D. degree from the University of Toronto, Toronto, Canada, in 2009, all in electrical engineering. In 2010, he joined the Department of Electrical and Electronic Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, where he is currently an Associate Professor. His current research interests include machine learning, computer vision, signal processing, and their applications.

Dr. Danilo Braga is a distinguished mechanical engineer and vibration specialist, currently serving as a Specialist at Dynamox SA. With a strong application background and a Ph.D. degree in noise and vibration, Danilo possesses extensive expertise in vibration monitoring and analysis, wireless sensor technologies, and signal processing. Also, Danilo has played a pivotal role in software development for vibration analysis (Dynamox's platform), contributing to the creation of innovative solutions that streamline the monitoring process and enhance fault prediction and diagnosis.